Is there a Supermassive Black Hole in the Nucleus of M31?

NASA Grant NAG5-2673

Annual Reports Nos. 1 and 2

For the Period 15 August 1994 through 14 August 1995
and 15 August 1995 through 14 August 1996

Principal Investigator
Dr. Francis A. Primini

September 1996

Prepared for:

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771

Smithsonian Institution
Astrophysical Observatory
Cambridge, Massachusetts 02138

> The Smithsonian Astrophysical Observatory
> is a member of the
> Harvard-Smithsonian Center for Astrophysics

Work during the report period focused on four areas of study, briefly described below:

1. Coordinated x-ray/radio monitoring of the nucleus of M31.

   A simple model for a spherically accreting super massive black hole suggests that it would be both an x-ray and a radio source, and that its x-ray and radio intensities would be anti-correlated. In collaboration with radio astronomers using the VLA, we have obtained near-simultaneous x-ray and radio observations of the nucleus at two epochs in July 1994. Preliminary results suggesting the presence of an anti-correlation were presented at the 1995 Tucson AAS meeting. However, this is part of a long-term study, and additional coordinated x-ray/radio observations are required before one can confidently claim a result. We have obtained additional observations in January of 1996 and are currently awaiting the results of the VLA analysis.

2. Precise determination of the position of the nuclear x-ray source.

   The VLA position of the nucleus is at the edge of the position error circle for the x-ray source. The source density in the region is sufficiently large that a effort to determine the x-ray position more carefully is justified. We have begun an effort to re-measure the x-ray position of the nucleus using Einstein HRI observations made in 1979 and ROSAT observations made in 1990, 1994, 1995, and 1996. We have used the x-ray detected globular clusters to do the astrometry, and the same analysis procedures throughout. Final analysis is underway and we expect to submit a paper in the near future.

3. Characterization of variability for all sources in the field of view.

   The center of M31 has been observed with good sensitivity on 3 occasions during the Einstein era and three occasions (1990, 1994, and 1996) during the ROSAT era. We are compiling a grand list of all the x-ray sources observed and a history of their intensities. We hope to provide a better estimate of the total number of sources in M31 and an estimate of the rate of transients. Preliminary results up to and including 1994 observations were presented at the 1995 Tucson AAS meeting.

4. Development of a technique for determination of intensities of point sources in a crowded x-ray field.

   The high density of sources near the nucleus precludes the use of simple synthetic aperture photometry to accurately determine their intensities. We have developed and coded a technique for simultaneous determination of the intensities of x-ray sources with overlapping point response functions. The technique is described in a paper which has been submitted to the Astrophysical Journal. A copy of paper is attached.

An Iterative χ^2 Minimization Technique for Bias-Free Parameter Estimation with Few Events per Bin

F.A. Primini and K. Kearns[1]

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138

ABSTRACT

We present a modified χ^2 fitting technique appropriate for fitting models to binned data with few events per bin. We demonstrate through numerical simulations that model parameters estimated with our technique are essentially bias-free, even when the average number of events per bin is ~ 1. This is in contrast to the results from traditional χ^2 techniques, which exhibit significant biases in such cases. The technique is relatively simple to use and can be easily incorporated into existing parameter-fitting programs.

1. Introduction

X-ray and γ-ray astronomers are often faced with the problem of extracting physical parameters from observations consisting of a small number of discrete photon "events". Often, this is accomplished by combining the events into some number of "bins", defined by the independent variables of the problem, and adjusting the parameters of a model until the predicted values in those bins compare well with observed values. The comparison is quantified through a test statistic, whose extremum defines the model parameters which best fit the data. If the number of events per bin is sufficiently large that the probability distribution in each bin can be assumed to be Gaussian, a good estimate of model parameters is provided by minimization of the χ^2 statistic, defined here in general terms as

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - P_i(\alpha_1, \alpha_2...\alpha_k))^2}{P_i(\alpha_1, \alpha_2...\alpha_k)}, \tag{1}$$

where O_i represents the observed number of events in the ith bin and $P_i(\alpha_1, \alpha_2...\alpha_k)$ the number predicted from a model with k parameters $\alpha_1, \alpha_2...\alpha_k$ (cf. Cramér 1946). Use of the χ^2 statistic offers the additional advantage that the probability distribution of χ^2 itself is well-known, and hence, goodness-of-fit can be easily assessed. However, when the number of events per bin is small, parameter estimates derived from χ^2 minimization can be significantly biased. Often, this is due to the commonly-used approximation of replacing P_i with O_i in the denominator of eq. 1 (Bevington and Robinson 1992, Wheaton et al. 1995, Churazov et al. 1995), but biased parameter estimates may also be obtained without making this approximation (Nousek and Shue 1989).

---

[1]present address: Dept. of Astronomy, Wesleyan University, Middleton, CT 06459

The cause of these biases is sometimes attributed to the use of a statistic associated with the Gaussian probability distribution in cases in which the Poisson distribution clearly applies. Indeed, for linear models, maximum-likelihood test statistics such as the C statistic (Cash 1979), which are based directly on the likelihood function of observing a particular set of events assuming an underlying Poisson distribution, yield unbiased parameter estimates in the few event limit. However, as Wheaton et al. (1995) and Churazov et al. (1995) have recently shown, unbiased parameter estimates are also possible using the χ^2 test statistic, provided the proper approximations are made for P_i in the denominator of eq. 1. In fact, Wheaton et al. show that if one replaces the P_i with the (unknown) expectation values $E_i = E[O_i]$, both the χ^2 and maximum-likelihood treatments reduce to the solution of the same set of equations. Of course, since the E_i depend on the unknown answers, they must be estimated in the minimization process, and both Wheaton et al. (1995) and Churazov et al. (1995) present techniques for doing so, in the context of the problems they wish to address.

In this paper, we present a particularly simple and general technique for estimating the E_i, and demonstrate through numerical simulations the validity of the technique in two simple problems of interest to X-ray astronomers. We begin, in the next section, with an examination of a simple χ^2 minimization problem. Our intent is to provide some insight into the biases inherent in traditional χ^2 minimization and to give a motivation for our technique, which we present in detail in section 3. We present the results of our numerical experiments in sections 4 and 5. We conclude with a brief discussion of reasons for considering the χ^2 statistic even when other test statistics are available.

2. Biases in Traditional χ^2 Minimization

We consider the trivial problem of estimating the sample mean, m, for randomly-occurring events with true mean μ events per bin, from a set of observed number of events, O_i, in N equally-sized bins. Eq. 1 can then be rewritten as

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - m)^2}{m} \tag{2}$$

which may be minimized analytically by solving $d\chi^2/dm = 0$ for m, yielding

$$2 \sum_{i=1}^{N} \frac{(O_i - m)}{m} + \sum_{i=1}^{N} \frac{(O_i - m)^2}{m^2} = 0. \tag{3}$$

The second term in eq. 3 reflects the bias in the fitting technique, since, without it, one would have the familiar result

$$m = \frac{1}{N} \sum_{i=1}^{N} O_i \tag{4}$$

with expectation value $E[m] = \mu$. The solution to eq. 3, however, is

$$m' = \left(\frac{1}{N}\sum_{i=1}^{N} O_i^2\right)^{1/2},\tag{5}$$

with expectation value $E[m'] \neq \mu$.

In the general case, χ^2 is minimized by solving the set of equations

$$\frac{\partial \chi^2}{\partial \alpha_k} = -2\sum_{i=1}^{N}\left(\frac{(O_i - P_i)}{P_i} + \frac{(O_i - P_i)^2}{2P_i^2}\right)\frac{\partial P_i}{\partial \alpha_k} = 0.\tag{6}$$

Generalizing on the results of equations 4 and 5, it seems reasonable to expect a biased result due to the second term in eq. 6. We argue that this is indeed the case by demonstrating that, in the absence of that term, eq. 6 reduces to a set of equations identical to those derived from the maximum likelihood method, which does produce unbiased parameter estimates for linear models. Following Cramér (1946), we write $P_i = np_i$, where n is the total number of events and p_i the probability of an obtaining an event in the ith bin. Eq. 6 then reduces to

$$\sum_{i=1}^{N}\frac{O_i}{p_i}\frac{\partial p_i}{\partial \alpha_k} = 0,\tag{7}$$

since it is assumed that $\sum p_i = 1$. With the likelihood function defined as

$$\mathcal{L} = \prod_{i}^{N} p_i^{O_i},\tag{8}$$

the maximum likelihood parameter estimates are given by the solution to the set of equations

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = \sum_{i=1}^{N} O_i p_i^{O_i-1}\frac{\partial p_i}{\partial \alpha_k}\prod_{j\neq i} p_j^{O_j} = \sum_{i=1}^{N}\frac{O_i}{p_i}\frac{\partial p_i}{\partial \alpha_k}\mathcal{L} = 0.\tag{9}$$

Under the assumption that $\mathcal{L} > 0$, this is equivalent to eq. 7.

We do not offer the above argument as a general proof, since there may be a number of unbiased estimators for a given problem. We only wish to provide some motivation for what follows. Moreover, since maximum-likelihood estimators are not, in general, unbiased, the claim of unbiased results through association with the maximum-likelihood formalism is, strictly speaking, only valid for those models which are known to have unbiased maximum-likelihood estimators, i.e., for linear models. However, as we shall see below, the technique does apply to some non-linear models of interest to x-ray astronomers.

3. A Technique for Unbiased Parameter Estimates

To avoid the biases demonstrated in the previous section, we seek estimates for the P_i in the denominator of eq. 1, which may be treated as constants in differentiation with respect to the α_k.

Ideally, we would like to use values derived from the parent distributions for each bin, but, of course, these are unknown a priori. However, if the model being fit provides a reasonable representation of the data, we may obtain useful estimates of the P_i through an iterative procedure. We rewrite eq. 1 as

$$\chi_j^2 = \sum_{i=1}^{N} \frac{(O_i - P_i^j)^2}{P_i^{j-1}}, \tag{10}$$

where j is an iteration index. For our present purposes, we restrict the technique to fitting for the total number of events per bin, including possible background events. The model is then guaranteed to be positive in each bin, and empty bins will not introduce any singularities in the calculation of χ^2.

We proceed as follows. At each iteration step, j, the P_i^{j-1} in the denominator of eq. 10 are approximated by the values of the model, using the best-fit parameters $(\alpha_1^{j-1}, \alpha_2^{j-1}...)$ of the previous iteration. These estimates depend on neither the O_i nor the current model P_i^j, being completely fixed by the results of the previous iteration. χ_j^2 is then minimized by standard techniques to produce a new set of best-fit parameters $(\alpha_1^j, \alpha_2^j...)$, which are in turn used to estimate the P_i^{j-1} for the subsequent iteration. For the initial iteration, $P_i^0 = 1$ for all i. The process may be terminated when differences in either χ^2 or the best-fit parameters fall below a suitable threshold. In our experience, in a number of different problems, convergence is typically achieved in ~5 iterations.

We emphasize that once the P_i^{j-1} have been determined from the results of the previous iteration, the solution for the new best-fit parameters for the current iteration can be treated as a standard χ^2 minimization problem. No special techniques are required[2]. The technique may thus be incorporated into existing parameter-fitting codes by the addition of a single, simple iteration loop. Although the need to repeat the minimization process a number of times does impose an additional computing burden, it is expected to be slight in most cases, given the power of modern computers.

4.   Experiment 1: Fitting a Simple Power Law

To demonstrate our technique, we repeat the numerical experiment of Nousek and Shue (1989), and simulate pulse height spectra for a simple power law with photon spectral index $\gamma = 2$, from 0.095 to 0.845 keV. For a range of total counts, N, from 25 to 1000, the true pulse-height spectra are given by

$$\overline{n}_i = N_o \int_{E_i}^{E_i + \Delta E} E^{-2} dE, \quad i = 1,...15, \tag{11}$$

---

[2]In fact, the problem may have become simplified. Since the denominator in the χ^2 sum no longer depends on the current parameters, non-linear minimization techniques may no longer be necessary. This, of course, depends on the nature of the particular model.

where

$$E_i = 0.095 + (i - 1)\Delta E = 0.095 + (i - 1)\frac{0.845 - 0.095}{15}, \tag{12}$$

and

$$N = N_o \int_{0.095}^{0.845} E^{-2} dE. \tag{13}$$

For each true spectrum, we simulated 1000 sample spectra $\{n_i\}$, where each n_i was a random deviate drawn from a Poisson distribution with mean $\bar{n}_i$. To generate Poisson deviates we used the Press routine poidev.c (Press et al. 1988), modified to use the random number generator ran0 rather that ran1. For each sample spectrum, we determined the best-fit parameters N_o^{fit} and γ^{fit}, using the technique described in section 3 and Powell's Method for function minimization (Press et al. 1988). This is the same minimization technique used by Nousek and Shue (1989). For each fit, parameter initial guesses were set to 50% of the true values, and conjugate directions were initialized to unit vectors.

In Table 1, we show the ratios of the average best-fit parameters to the true values for a number of different values of total counts N. For $N = 25$, ~98% of the fits converged. For all other values of N, virtually all fits ($\gtrsim$99%) converged. For comparison, we also show the corresponding results from Table 3 of Nousek and Shue (1989), where fitting is accomplished either by C-statistic minimization or standard χ^2 minimization. For all values of N tested, the results of the iterative χ^2 minimization technique exhibit little or no bias, and are comparable in accuracy to the results of the C-statistic minimization.

To assess the utility of our technique in estimating confidence levels, we calculated the value of χ^2 corresponding to the true parameters for each sample spectrum. We then determined the percentage of the fits for which the $\chi^2_{min} + \Delta\chi^2$ contours included χ^2_{true} for $\Delta\chi^2$ values appropriate to various joint two-parameter confidence levels. Our results are shown in Table 2 and indicate that iterative χ^2 minimization provides reasonable estimates for standard confidence levels, even for small N.

Finally, we compare our sample distributions of χ^2_{min} with the theoretical χ^2 distribution. Our results are shown in Table 3 and in Figure 1. One-sample Kolmogorov-Smirnov tests indicate that for $N > 50$, the sample distributions are consistent with the theoretical distribution at the 99% confidence level.

5. Experiment 2: Fitting Source and Background Intensities to Image Data

As a second example of iterative χ^2 minimization, we consider the problem of determining the intensities of unresolved x-ray sources at known locations in an x-ray image, in the presence of a uniform background. Specifically, we consider the case where the separation between sources is sufficiently small that the individual source photon distributions overlap significantly. In this case, simple synthetic aperture photometry techniques prove inadequate, since each source aperture will

contain an unknown number of events from other nearby sources. Rather, the intensities of all sources whose photon distributions overlap must be fit simultaneously. In general, the predicted number of events in image pixel ij may be written

$$P_{ij} = B + \sum_{k=1}^{N} T_{ijk} S_k, \tag{14}$$

where B represents the number of background events per pixel, S_k is the total number of events due to the kth source, and T_{ijk} is the integral of the point response function, centered at the location of the kth source, over the area of pixel ij. Assuming the T_{ijk} are known, the values of B and S_k may be fit by iterative χ^2 minimization, where eq. 10 is replaced by

$$\chi_l^2 = \sum_{i,j} \frac{(O_{ij} - P_{ij}^l)^2}{P_{ij}^{l-1}}, \tag{15}$$

and the sum is now over all pixels ij of interest.

We again use simulations to demonstrate the technique. We assume an ideal 256×256 pixel image containing two sources near the center, separated by 10 pixels. We use the ROSAT HRI on-axis point response function (David et al. 1994) to describe the source photon distributions, and assume source intensities of $S_1 = 250$ and $S_2 = 50$ events each. For each of four values of B, corresponding to 500, 1000, 5000, and 10000 total background events, we have simulated 1000 sample images, where the value, O_{ij}, of each image pixel is a random deviate drawn from a Poisson distribution with mean $\bar{n}_{ij}$ given by eq. 14. The maximum number of events per pixel is typically $\sim$5, near the peak of the brighter source, and the average number of events per background pixel ranges from $\sim$0.008 for the low background case to $\sim$0.15 for the high background case.

For each sample image, we fit simultaneously for B, S_1, and S_2. For this particular problem, since the model is a linear function of the parameters, the minimum of χ^2 may be determined analytically. We use the method of LU decomposition with iterative improvement (Press et al. 1988). Our results are shown in Table 4, where again we tabulate the ratios of the average best-fit parameters to the true values. In all cases, the averages accurately reproduce the true values.

6. Conclusions

We conclude that biased parameter estimates arising from χ^2 minimization in the few event limit may be significantly reduced or eliminated by the use of the iterative technique for calculating χ^2 described here. Moreover, confidence levels can still be determined by standard $\chi_{min}^2 + \Delta\chi^2$ techniques, and in many cases the modified test statistic follows a distribution consistent with the χ^2 distribution. The technique is relatively simple to use, and, at each iteration step, reduces to a standard χ^2 minimization problem requiring no special techniques for solution. Therefore, it may be easily incorporated into existing parameter-fitting codes which employ χ^2 minimization.

It remains to address the question of why one should consider χ^2 -fitting at all in cases where maximum-likelihood techniques are effective. We offer four possible reasons. First, goodness-of-fit can, in general, be easily assessed using the χ^2 test statistic, whereas it cannot, using maximum-likelihood statistics, without extensive simulations. Second, we believe that it is always useful to have more than one tool which can be used to attack a particular problem, to allow the researcher some flexibility in the analysis and a means of cross-checking the results. Third, the astronomical community at large is at present far more familiar with χ^2 minimization techniques, and a χ^2 - fitting tool which can provide accurate results in the few event limit can be useful in testing the development of maximum-likelihood fitting techniques. Finally, there may be situations in which the researcher cannot know a priori that the few event limit has been reached. This is likely to occur, for example, in the automated analysis of a large number of datasets, in which the number of events per dataset can span a wide range. It may then prove difficult, or at least extremely inconvenient, to determine the appropriate fitting technique on a case-by-case basis. Rather, one would prefer to use a single fitting technique which can produce bias-free results for all datasets. Should maximum-likelihood fitting prove too costly for datasets with many events, a technique such as iterative χ^2 minimization may be preferable. In the end, however, we cannot, in general, advocate one fitting technique over the other. That choice must be made by the individual researcher based on the details of the problem to be solved.

Table 1. Comparison of best-fit parameters from iterative χ^2 minimization and other fitting techniques.

N	Iterative χ^2 Minimization		C-statistic Minimization		Standard χ^2 Minimization	
	N_o^{fit}/N_o^{true}	$\gamma^{fit}/\gamma^{true}$	N_o^{fit}/N_o^{true}	$\gamma^{fit}/\gamma^{true}$	N_o^{fit}/N_o^{true}	$\gamma^{fit}/\gamma^{true}$
25	1.145	1.003	1.269	0.958	0.709	1.152
50	1.055	1.008	1.079	0.998	0.647	1.134
75	1.025	1.009	1.078	0.995	0.636	1.130
100	1.008	1.008	1.053	0.996	0.673	1.109
150	1.025	1.001	1.015	1.005	0.707	1.094
250	1.019	1.000	1.019	1.000	0.767	1.072
500	1.007	1.000	0.997	1.004	0.863	1.040
1000	1.005	1.000	1.001	0.999	0.937	1.017

Table 2. Percentage of fits for which $\chi^2 min + \Delta\chi^2$ includes χ^2_{true}.

N	$\Delta\chi^2$ (Expected Percentage)					
	2.30 (68.3%)	4.61 (90%)	6.17 (95.4%)	9.21 (99%)	11.8 (99.73%)	18.4 (99.99%)
25	69.8	87.0	92.1	96.8	98.4	99.9
50	68.9	88.1	93.7	97.2	98.2	98.3
75	67.7	87.8	93.5	98.4	99.3	99.8
100	68.1	89.1	94.1	98.2	99.0	99.8
150	67.0	87.1	93.6	97.7	99.3	100
250	68.0	90.3	95.6	99.2	99.8	100
500	69.1	90.1	95.5	98.9	99.6	99.9
1000	69.6	88.9	95.0	99.0	99.7	100

Table 3. Results of One-Sample K-S Tests Comparing Sample and Theoretical χ^2 Distributions.

N	$P(D > D_{max})$
25	0.004
50	0.002
75	0.047
100	0.998
150	0.894
250	0.290
500	0.812
1000	0.958

Table 4. Results of fits to background and source intensities in images.

Background Events	B^{fit}/B^{true}	S_1^{fit}/S_1^{true}	S_2^{fit}/S_2^{true}
500	0.999	0.997	0.995
1000	0.999	0.999	1.004
5000	1.001	0.997	0.999
10000	1.000	1.000	1.002

REFERENCES

Bevington, P.R. and Robinson, D.K. 1992, Data Reduction and Error Analysis for the Physical Sciences, Second Edition (New York: McGraw-Hill)

Cash, W. 1979, ApJ, 228, 939

Churazov, E., Gilfanov, M., Forman, W. and Jones, C. 1995, submitted to ApJ

Cramér, H. 1946, Mathematical Methods of Statistics (Princeton: Princeton University Press)

David, L.P. et al. 1994, ROSAT HRI Final Calibration Report

Nousek, J.A. and Shue, D.R. 1989, ApJ, 342, 1207

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vettering, W.T. 1988, Numerical Recipes in C (Cambridge: Cambridge University Press)

Wheaton, W.A., Dunklee, A.L., Jacobson, A.S., Ling, J.C., Mahoney, W.A. and Radocinski, R.G. 1995, ApJ, 438, 322

Fig. 1.— Comparison of sample and theoretical χ^2 distributions for Experiment 1. Histograms of χ^2_{min} values, normalized to unit area, and sample cumulative distributions are indicated by solid curves. Probability densities and cumulative distributions for the χ^2 distribution sre indicated by dotted curves. a) $N = 25$. b) $N = 100$.